

FINAL  
1995/96 Progress Report for NAGW-2108

## A: BACKGROUND

It is generally believed that an RNA World existed at an early stage in the history of life. During this early period, RNA molecules are seen to be potentially involved in both catalysis and the storage of genetic information. It is widely believed that this RNA World was extensive and therefore a sophisticated nucleic acid replication machinery would presumably predate the translation machinery which would not be needed until later stages in the development of life. This view of an extended RNA World is not necessarily correct. One might alternatively envision (Smith & Fox, 1995; Microbiologia SEM 11: 217-224, 1995), an abbreviated RNA World where peptide synthesis commenced very early and translation essentially coevolved with replication. From the point of view of exobiology, the difference in these two views mainly affects the significance of studies of the extent of catalysis possible by RNA. In either case, the origin of the translation machinery and the principles of RNA evolution remain central problems in exobiology.

Translation presents several interrelated themes of inquiry for exobiology. First, it is essential, for understanding the very origin of life, how peptides and eventually proteins might have come to be made on the early Earth in a template directed manner. Second, it is necessary to understand how a machinery of similar complexity to that found in the ribosomes of modern organisms came to exist by the time of the last common ancestor (as detected by 16S rRNA sequence studies). Third, the ribosomal RNAs themselves likely had a very early origin and studies of their history may be very informative about the nature of the RNA World. Moreover, studies of these RNAs will contribute to a better understanding of the potential roles of RNA in early evolution.

The actual history of translation appears very complex. The problem is accentuated by the fact that a major portion of that history likely took place during a transition period before the final emergence of the last common ancestor as defined by the 16S rRNA phylogeny. Indeed, it is not unreasonable to suppose that it was obtainment of a sufficiently proficient translation machinery that was the final and decisive step in the emergence of true life. To date, the vast majority of work on translation has focused on function rather than the historical origins of the machinery. The work which has been conducted has mostly focused on individual components and the larger picture has languished. It is the goal of his project to begin to address this history from the overall perspective. The approach being used is multifaceted. First, we are using informatics to study what can be learned about the history of translation by the comparative studies of genomes. Second, direct experimental studies of rRNA evolution and the possible co-evolution of rRNA and ribosomal proteins are being undertaken. This work employs 5S rRNA as a model system. The final component of the project is a direct, but highly speculative, assault on a key historical question: Can simple RNA molecules that might have arisen in an RNA world participate in peptide synthesis if they are carrying amino acids? Success here would link proposals about the late prebiotic world with the eventual modern translation system.

## B: PROGRESS - 12/1/95 - 1/30/96

During the past year we have conducted a comparative study of four completely sequenced bacterial genomes. We have focused initially on conservation of gene order. As we reported at the ISSOL meeting in July, that gene order is remarkably unconserved. Nevertheless, the genes whose order is conserved include essentially all the components of the translation machinery. This strongly reinforces our belief that the ribosomal machinery came into existence at a very early time. Since conservation of gene order almost certainly reflects conservation of coordinated gene expression, this list of conserved gene orders, Table 1, likely reflects the oldest regulatory elements in the cell. Hence, since many of the ribosomal protein gene orders are also conserved in the Archaea genome, it is clear that not only did these genes exist very early, but their expression was already coordinated to at least some extent prior to the emergence of the last common ancestor as defined by 16S rRNA studies. How are these genes regulated? Although this type of data is typically only available in *E. coli*, the result is impressive. It turns out that in every case where it is known, the conserved gene clusters are regulated by RNA mechanisms rather than DNA level mechanisms. We believe this is strong evidence for the existence of an RNA genome prior to the modern DNA genome. We are currently preparing a paper reporting these findings and ideas that will be submitted to *Nature* in early December of 1996.

The second component of the project continues to build on the model system for studying the validity of variant 5S rRNA sequences in the vicinity of the modern *Vibrio proteolyticus* 5S rRNA that we established earlier (Hedenstierna *et al.*, Syst. Appl. Microbiol. 16: 280-286 (1993) and Lee *et al.*, Origins Life & Evol. Biosphere 23: 365-372 (1993)). This system has made it possible to conduct a detailed and extensive analysis of a local portion of the sequence space. This system was subsequently extended to include a deletion construct to which replacement sequence segments can be readily added (Pitulle *et al.*, Appl. Environ. Microbiol. 61: 3661-3666 (1995) and Pitulle *et al.*, (1997)). As it exists, this system allows us to explore a typical RNA sequence space, accessing both validity and invalidity of various sequences. It is our goal to be able to enhance current parsimony rules for constructing likely historical sequences from analysis of extant sequences for RNA. One long term hope would in fact be the actual synthesis and testing of ancient sequences as has been pioneered by Benner for proteins. We have already used this approach to reconstruct all possible equally parsimonious evolutionary intermediates between two pairs of sequences in RNA sequence space (Lee *et al.*, manuscript submitted to J. Molecular Evolution) and show that not all apparently equally likely paths are in fact equally probable.

These core methods have been used to construct numerous mutants during the last several years. Although it has been a secondary focus, this work has continued over the last year such that we now have in excess of 125 *V. proteolyticus* derived constructs which have been made and characterized. These include 74 of the possible 360 point mutations. The vast majority of these constructs exhibit one of three major phenotypes. Type 1

**Table I**

**Function of Gene or Protein in *E. coli*/S. typhurium**

Group	Short Name, Gene Names	Function of Gene or Protein in <i>E. coli</i> /S. typhurium
I	<p>16s, 23s, 5s rRNAs (16S, 23S &amp; 5S rRNA)</p> <p>S10 "operon" (<i>rps10, rpl3, rpl4, rpl23, rps19, rpl22, rps3, rpl16, rpl29, rps17</i>)<sup>a</sup></p> <p>Str "operon" (<i>rps7, fuxG</i>)<sup>a</sup></p> <p>Spc "operon" (<i>rpl14, rpl24, rpl3, rps14, rps8, rpl6, rps18, rps5, rpl30, rpl15, SecY</i>)<sup>a</sup></p> <p>L13 "operon" (<i>rps9, rpl13</i>)</p> <p>L11 "operon" (<i>rpl11, rpl1</i>)<sup>a</sup></p> <p>Alpha "operon" (<i>rps13, rps11, rps4, rpl17</i>)<sup>a</sup></p> <p>L35 "operon" (<i>infC, rpl35, rpl20</i>)<sup>a</sup></p> <p>L34 "operon" (<i>rnp4, rpl34</i>)</p> <p>L21/L27 (<i>rpl21, rpl27</i>)</p> <p>L10 "operon" (<i>rpl10, rpl12</i>)</p> <p>S6 "operon" (<i>rps6, rps18, rpl9</i>)</p> <p>tmD "operon" (<i>rpl19, rmd, rps16</i>)</p>	RNA and protein constituents of the ribosome
II	<p><i>ATP<sub>max</sub></i> (<i>atpC, atpD, atpG, atpA, atpH, atpF, atpE, atpB</i>)<sup>a</sup></p>	<p>ATP synthetase, codes for hydrophilic F<sub>1</sub> unit and a membrane bound hydrophobic F<sub>0</sub> unit.</p> <p>Forms proton channel</p>
III	<p>Beta "operon" (RNA Polymerase) (<i>rpoC, rpoB</i>)<sup>a</sup></p> <p>Initiation factor (<i>musA, infB</i>)<sup>a</sup></p> <p>Preprotein translocase/transcript. antiterm (<i>secE, musG</i>)</p>	<p>B and B' subunit of bacterial RNA polymerase, (A' and A" subunits in Archae)</p> <p>Modulates rate of chain synthesis, can prolong pausing at natural pause sites. Binds special tRNA to the ribosome complex.</p> <p>Transmembrane component, minimal domain conserved in bacteria and Eukaryotes point to important role in protein export. Enhances rate of transcription elongation.</p>
IV	Spermidine/putrescine Transport ( <i>potA, potB, potC</i> )	Transporter of poly amines into cell
V	Chaperones* ( <i>groEL, groES</i> )	HSP, binds to wide range of nonnative polypeptide substrates and nascent chain RNA for proper folding and reduction of aggregation.
VI	Pyruvate dehydrogenase ( <i>pdhC, pdhE</i> )	Responsible for production of CoA from pyruvate. Coenzyme A is universal carrier of acyl groups.

constructs exhibit essentially wild type behavior and thus appear to be valid 5S rRNAs. Type 2 constructs do not accumulate to substantial levels in the cell. This apparently reflects the instability of the product as a result of either overprocessing originally or, less likely, premature degradation. These variants are found primarily in the region of the molecule that has been most strongly implicated in the binding of ribosomal protein L18. Type 3 constructs accumulate to very high levels but are absent from both 50S ribosomal subunits and 70S ribosomes. Type 3 mutations are especially common in the helix III/loop C subdomain of 5S rRNA which also is the region of greatest sequence conservation.

The existing mutants raise several questions which can not be answered without knowledge of the interaction of 5S rRNA with its proteins, especially L18. Thus for example, is loss of stability (type 2 phenotype) a consequence of loss of protein binding ability? Is the loss of ability to incorporate related to protein binding? During the past year we have continued our efforts begun two years ago (Setterquist *et al.*, Gene, in press) to bring the protein component into our analysis. During the past year we focused on binding studies in which we used filter binding assays to determine the affinity of L18 to various mutants. It has now been shown (Lee *et al.*, manuscript in preparation), that RNA stability is not directly linked to L18 affinity to the RNA. Our level of interest in the RNA/protein interaction has increased as we have begun to appreciate that RNA binding proteins are in effect exhibiting RNA chaperonin activity. Their interaction with the RNA changes the RNA's structure and allows it to fold in such a manner that it is functional. Thus, RNA/protein interactions may have been essential not only to the structure of the early ribosome but to stabilization of the early genome itself. The early emergence of r-proteins discussed in the first section is quite consistent with this. Moreover, there is evidence that the various ribosomal proteins may be historically related and likely share common motifs of interaction with the RNA. Comparative studies of several rRNA/protein binding sites are already underway (as a community effort by different groups working on different rRNA/r-protein pairs). The 5S rRNA/L18 interaction is one of the better studied and thus our work is important in this context too.

We have also continued high resolution NMR work on RNA oligomers originally initiated by G. Kenneth Smith who was funded by a NASA Graduate Student Researcher's Fellowship Award until May of 1996. Mr. Smith developed synthesis & purification protocols in order to obtain large quantities of RNA oligomers for NMR studies. He succeeded in obtaining preliminary one dimensional spectra for an analog of the helix II/loop B/helix III domain of 5S rRNAs well as a 29mer which includes helix III and the highly characteristic 13 nucleotide loop C. The downfield spectra of the loop C domain suggests several nonstandard interactions may be present. Such interactions would likely provide a unique three dimensional folding that would not easily be duplicated. This would explain the large number of Type 3 mutants in loop C. The preliminary results suggest that both of these structures can be solved at near atomic resolution. More recently, a new student has synthesized a large quantity of the 29mer and succeeded in obtaining 2D spectra in both H<sub>2</sub>O and D<sub>2</sub>O. She is now attempting to do assignments and it is anticipated that enough insight will be obtained for a preliminary publication. In addition,

we have obtained an excellent preliminary spectra of L18 and it should also be possible to study this molecule's structure in solution at atomic resolution. During the previous year, we discovered that a 63 nucleotide double deletion mutant in which almost all of helix III, loop C and helix IV have been deleted still exhibits the Type 3 phenotype. During the past year, we conducted protein binding studies on this RNA and found that when it is produced by T7 runoff transcription that it can bind L18 with essentially wild type affinity. These various NMR results suggest that considerable high resolution structure work can be done on both the 5S rRNA and the RNA/protein complex. Although these results will be of value to exobiology, our perception is that they will be of greater significance in basic science. It therefore is our intention to seek grant support from other agencies for full scale continuation of this work.

One of the major limitations of our approach to RNA evolution has been the rate at which the RNA sequence space can be explored. In alternate approaches such as *in vitro* RNA evolution as practiced by Joyce, Szostak and others, large numbers of sequences can be tested. Although this approach is very fast, it does not fully meet our needs as it is (a) *in vitro* and (b) does not guarantee that any particular sequence was tested - i.e. it finds what works for a particular selection but does not readily identify what does not work. During the past year we have begun development of what we believe will be a major improvement on our system that will allow testing of a large number of mutants simultaneously. The idea is this: (1) Eliminate some or all of the wild type genomic 5S rRNA genes in the test strain such that the strain is either extremely sick or even dead (2) restore cell viability using a compensatory plasmid which carries a valid 5S rRNA gene; (3) destroy the cells' recombination machinery so that it can not incorporate the plasmid into its main genome. This strain can then be used to simultaneously test the validity of large numbers of 5S rRNA variants *in vivo* by transforming cells cured of the plasmid with a mixture of potentially compensatory plasmids. Those cells that grow will necessarily contain a plasmid with a valid 5S rRNA. The identity of the valid 5S rRNA can then be rapidly identified by sequencing. During the past year we have invested considerable technical effort on construction of the cell line needed for this approach. So far we believe we have succeeded in knocking out four of the 8 genomic 5S rRNA genes. The current cell line appears to be very sick and we have managed to restore growth with a plasmid carrying a 5S rRNA gene. It therefore appears that this new approach is feasible.

The third phase of our effort is a new and highly speculative component of our effort that was facilitated by two events; (1) A visit of Ms. Lisa DSouza to Dr. David Deamer's laboratory as a Planetary Biology Intern where she learned RNA encapsulation techniques and (2) the part time association with my laboratory of Dr. Susan Martinis who is expert in tRNA synthetase and knows how to charge both normal tRNAs and minihelix analogs. The goal of our effort is to attempt to demonstrate that conditions exist under which charged tRNA minihelices can participate in peptide bond formation in the absence of modern ribosomes. The synthesis of a peptide bond is thermodynamically favorable and hence the problem may simply be one of orientation. Success in this effort would be an important demonstration that one could have primitive template directed peptide

synthesis in the absence of a full fledged RNA World. Our plan of action is to charge leucine minihelices and attempt to identify conditions that result in peptide bond formation. Regardless of whether or not the peptidyl transferase reaction utilizes chemical catalysis, it is clear that proper orientation of the aminoacylated RNA minihelices, such that the aminoacylated ends of the molecule will be in close proximity, will be important. This may be achieved in a variety of ways. We intent to accomplish this by the dehydration/rehydration lipid encapsulation procedure Lisa DSousa learned while in Dr. Deamer's laboratory. Vesicles produced in this way are generally permeable to small ions but should hold the minihelices quite easily. They will be able to transport leucine monomer and probably the hydrophobic leucine dipeptide.

Significant progress in laying the required groundwork for this work has been accomplished during the past year. A plasmid, pLeuS-1, carrying the *E. coli* leucine tRNA synthetase gene (LeuRS) has been obtained and used to overexpressed the leucine synthetase. A T7 runoff transcription has been established in our laboratory and successfully used to make both minihelix RNAs and tRNAs. In addition, encapsulation studies have been conducted and we have successfully encapsulated tRNA in lipid vesicles. Initial efforts to obtain peptide bond synthesis will be underway over the next several months.

#### **D: OUTLINE OF WORK PLANNED FOR COMING YEAR**

The work in the coming year will largely be a continuation of what has been begun in the previous year. In the theoretical component we will begin to assemble aligned sequence data sets of all ribosomal components in each of the completely sequenced genomes using the GDE multiple sequence alignment package. This will include all proteins and RNAs that are either directly (e.g. ribosomal proteins, factors, etc). or indirectly (methylases, rRNA processing enzymes, etc) associated with the ribosomal machinery. In doing this, we will initially concentrate on components that have been largely ignored while initially relying whenever possible on published analysis of components that have been extensively studied. We will however carefully analyze all the proteins that bind directly to the ribosomal RNA to detect either duplications or fusion events that may have occurred in their history. This is because these proteins are likely to have been key components in ribosome evolution. We will also examine noncoding regions of the various r-protein operons to see if the RNA level signals for translational control detected in many cases in *E. coli* are in fact phylogenetically conserved as would be predicted by our findings of the past year.

Our studies on RNA sequence space will primarily focus on two major efforts begun in the past year. These are (1) the further development and initial testing of the knockout/compensatory plasmid system for rapid exploration of 5S rRNA sequence space *in vivo* and (2) the reconstruction and testing *in vivo* of common ancestral sequences predicted by parsimony analysis of extant *Vibrio* 5S rRNA sequences. Several unambiguous hypothetical common ancestors that would have existed over 100 million

years ago have been identified. During the next year variant 5S rRNAs with these sequences will be constructed by successive mutagenesis and tested in our *in vivo* assay system for functionality. This work will be an important new test of the reasonableness of parsimony in predicting reasonable historical sequences. In addition, as time permits additional NMR work and RNA/protein interaction work will be continued while alternative funding for these efforts is sought. Finally, the first experiments will be undertaken to attempt to obtain ribosome independent peptide bond formation. This will initially be attempted with charged tRNAs using dipeptide catalysts and later with charged minihelices.

## E: PEER REVIEW PUBLICATIONS

Yang, Y. & Fox, G. E. "An Archaea 5S rRNA Analog is Stably Expressed in *Escherichia coli*" *Gene* **168**: 81-85 (1996).

Pitulle, C., Hedenstierna, K. O. F. & Fox, G. E. "Useful Properties of Restriction Enzymes that Recognize Interrupted Palindromes" *BioTechniques* **21**: 619-622 (1996).

Setterquist, R. A., Smith, G. K., Oakley, T. H. & Fox, G. E. "Sequence, Overproduction and Purification of *Vibrio proteolyticus* Ribosomal Protein L18 for *in vitro* and *in vivo* Studies", *Gene*, in press (1996).

Pitulle, C., DSouza, L., & Fox, G. E. "A Low Molecular Weight Artificial RNA of Unique Size with Multiple Probe Target Regions", *System. Appl. Microbiol.* **20**: in press (1996)..

Lee, Youn-Hyung Lee, DSouza, L. M. and Fox, G. E. "Equally Parsimonious Pathways Through an RNA Sequence Space are not Equally Likely", submitted to *J. Molecular Evolution*.

## F: PRESENTATIONS

Seifert, J. and Fox, G. E.; "The melding of morphology and molecular phylogeny"; A poster presented at the 96th General Meeting American Society of Microbiology; May 19-23, 1996; New Orleans, La.

Samartzidou, H., Abdi, F., Fox, G. E., Martin, K., and Widger, W. R.; "Sampled sequencing: A rapid method for developing a genetic map of *Synechococcus* PCC 7002"; A poster presented at the 96th General Meeting American Society of Microbiology; May 19-23, 1996; New Orleans, La.

Fox, G. E., Martin, K. A., Widger, W. R., Abdi, F., and Seifert, J. L.; "Bacterial operons containing components of the translation machinery are of great antiquity" Poster #129 presented at 11th International Conference on the Origin of Life, July 5-12, 1996, Orleans, France.

Seifert, J. L. and Fox, G. E.: "The melding of morphology and molecular phylogeny: Insight into ancestral prokaryotic phenotypes" Invited presentation at 11th International Conference on

the Origin of Life, July 5-12, 1996, Orleans, France.

Fox, G. E. "*In Vivo* Studies of a Local Region of 5S rRNA Sequence Space"; forthcoming seminar at Rice University; 12/9/1996 in Houston, Tx.

Fox, G. E. "Progenotes and Archaeobacteria: Facts and Hypotheses about Earliest Life"; forthcoming lecture in public lecture series entitled Origin of Life: Earth, Mars and Beyond" ; February 25, 1997; Houston, Texas.